Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs
and/or
Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)
May 2023
Biostatistics

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 2

III. RECOMMENDATIONS FOR COVARIATE ADJUSTMENT IN CLINICAL TRIALS 3

A. General Considerations 3

B. Linear Models 4

C. Nonlinear Models 5

IV. REFERENCES 8

Contains Nonbinding Recommendations

Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry[1]

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

I. INTRODUCTION

This guidance describes FDA's current recommendations regarding adjusting for covariates in the statistical analysis of randomized clinical trials in drug[2] development programs. This guidance provides recommendations for the use of covariates in the analysis of randomized, parallel group clinical trials that are applicable to both superiority trials and noninferiority trials. The main focus of the guidance is on the use of prognostic baseline covariates[3] to improve statistical efficiency for estimating and testing treatment effects. This guidance does not address use of covariates to control for confounding variables in non-randomized trials, the use of covariates in models to account for missing outcome data (National Research Council 2010), the use of covariate adjustment for analyzing longitudinal repeated measures data, the use of Bayesian methods for covariate adjustment, or the use of machine learning methods for covariate adjustment.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

[1] This guidance has been prepared by the Office of Biostatistics in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research at the Food and Drug Administration.

[2] The term *drug* used in this guidance refers to both human drugs and biological products.

[3] The term *prognostic baseline covariates* used in this guidance refers to baseline covariates that are likely to be associated with the primary endpoint. Use of *predictive baseline covariates* to identify groups more likely to benefit from treatment is beyond the scope of this guidance.

II. BACKGROUND

Baseline covariates in this guidance refer to demographic factors, disease characteristics, or other information collected from participants before the time of randomization. Covariate adjustment refers to the use of baseline covariate measurements for estimating and testing population-level treatment effects between randomized groups.

In many randomized controlled trials, the primary analysis used to estimate treatment effects of a new drug might not adjust for baseline covariates (through what is termed an unadjusted analysis). However, incorporating prognostic baseline covariates in the design and analysis of clinical trial data can result in a more efficient use of data to demonstrate and quantify the effects of treatment. Moreover, this can be done with minimal impact on bias or the Type I error rate.

The ICH guidance for industry *E9 Statistical Principles for Clinical Trials* (September 1998)[4] addresses these issues briefly. The ICH E9 guidance encourages the identification of "covariates and factors expected to have an important influence on the primary variable(s)." The ICH E9 guidance strongly advises prespecification of "the principal features of the eventual statistical analysis," including "how to account for [covariates] in the analysis to improve precision and to compensate for any lack of balance between treatment groups." The ICH E9 guidance also cautions against adjusting for "covariates measured after randomization because they could be affected by the treatments."

This guidance is consistent with the ICH guidance for industry *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials.* After specifying the treatment condition of interest, target population, and endpoint variable — the treatment effect estimated by covariate adjustment is a population summary measure defining an estimand.

This guidance provides general considerations and additional recommendations for covariate adjustment using linear and nonlinear models.[5] In linear models, adjustment for prognostic baseline covariates often leads to improved precision by reducing residual variance. When adjusting for covariates based on fitting nonlinear regression models, such as logistic regression models in studies with binary outcomes, there are additional considerations that arise because inclusion of baseline covariates in a regression model can change the treatment effect that is being estimated. As explained below, after suitably addressing the treatment effect definition, covariate adjustment using linear or nonlinear models can be used to improve statistical efficiency.

[4] We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/regulatory-information/search-fda-guidance-documents.

[5] For this guidance, nonlinear models can include generalized linear models with nonlinear link functions.

III. RECOMMENDATIONS FOR COVARIATE ADJUSTMENT IN CLINICAL TRIALS

A. General Considerations

- An unadjusted analysis is acceptable for the primary analysis of an efficacy endpoint.

- Sponsors can adjust for baseline covariates in the analyses of efficacy endpoints in randomized clinical trials. Doing so will generally reduce the variability of estimation of treatment effects and thus lead to narrower confidence intervals and more powerful hypothesis testing.

- Sponsors should prospectively specify the detailed procedures for executing covariate-adjusted analysis before any unblinding of comparative data. FDA review will emphasize the prespecified primary analysis rather than post-hoc analyses using different models or covariates.

- Covariate adjustment leads to efficiency gains when the covariates are prognostic for the outcome of interest in the trial. Therefore, FDA recommends that sponsors adjust for covariates that are anticipated to be most strongly associated with the outcome of interest. In some circumstances these covariates may be known from the scientific literature. In other cases, it may be useful to use previous studies (e.g., a Phase 2 trial) to select prognostic covariates or form prognostic indices.

- Covariate adjustment can still be performed with covariates that are not prognostic, but there may not be any gain in precision (or may be a loss in precision) compared with an unadjusted analysis.

- Covariate adjustment is acceptable even if baseline covariates are strongly associated with each other (e.g., body weight and body mass index). However, adjusting for less correlated baseline covariates generally provides greater efficiency gains.

- Randomization is often stratified by baseline covariates. A covariate adjustment model should generally include strata variables and can also include covariates not used for stratifying randomization. In some cases, incorrect stratification may occur and result in actual and as-randomized baseline strata variables. A covariate adjustment model can use either strata variable definition as long as this is prespecified.

- Sponsors can conduct randomization/permutation tests with covariate adjustment (Rosenbaum 2002).

- In a trial that uses covariate adjustment, the sample size and power calculations can be based on adjusted or unadjusted methods. The latter will often lead to a more conservative sample size.

- Clinical trials often record a baseline measurement of a defined characteristic and record a later measurement of the characteristic to be used as an outcome. Adjusting for the baseline value rather than (or in addition to) defining the primary endpoint as a change from baseline is generally acceptable. Sponsors proposing to define the outcome as a percentage change rather than an absolute change from baseline should discuss the outcome definition and use of covariate adjustment with the relevant review division. Sponsors proposing to perform noninferiority testing on ratios of treatment group means rather than differences of treatment group means should also discuss change from baseline outcome definitions and use of covariate adjustment with the relevant review division.

- Sponsors should discuss proposals for complex covariate-adaptive randomization, data-adaptive covariate selection, or use of covariate adjustment in an adaptive design with the relevant review division.

- The statistical properties of covariate adjustment are best understood when the number of covariates adjusted for in the study is small relative to the sample size (Tsiatis et al. 2008). Therefore, sponsors should discuss their proposal with the relevant review division if the number of covariates is large relative to the sample size or if proposing to adjust for a covariate with many levels (e.g., study site in a trial with many sites).

B. Linear Models

- Covariate adjustment through a linear model is an acceptable method for estimating the average treatment effect, which is the difference in expected outcomes between subjects assigned to treatment and control groups. Generally, the outcome is regressed on an intercept, treatment assignment indicator, and baseline covariates, and the model is estimated using ordinary least squares. The resulting estimated regression coefficient for the treatment indicator is the estimate of the average treatment effect.

- The average treatment effect is an example of an unconditional treatment effect, which quantifies the effect at the population level of moving a target population from untreated to treated.

- Even when the linear regression model is misspecified and does not accurately capture the relationships between the outcome, covariates, and treatment, covariate adjustment through a linear model is a valid method for estimating and performing inference for the average treatment effect (Lin, 2013). However, the power of hypothesis tests and precision of estimates generally improves if the model more closely approximates the true relationships among the outcome, covariates, and treatment.

- Nominal standard errors are often the default method in most statistical software packages. Even if the model is incorrectly specified, they are acceptable in two arm trials with 1:1 randomization. However, in other settings, these standard errors can be inaccurate when the model is misspecified. Therefore, the Agency recommends that sponsors consider use of a robust standard error method such as the Huber-White "sandwich" standard error when the model does not include treatment by covariate interactions (Rosenblum and van der Laan

2009; Lin 2013). Other robust standard error methods proposed in the literature can also cover cases with interactions (Ye et al. 2022). An appropriate nonparametric bootstrap procedure can also be used (Efron and Tibshirani 1993).

- An analysis ignoring stratified randomization is likely to overestimate standard errors and can be unduly conservative when performing inference for the average treatment effect. The Agency recommends that the standard error computation account for stratified randomization. There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (Bugni et al. 2018; Ye et al. 2021). The statistical properties of such methods are best understood when the number of strata is small relative to the sample size. Sponsors can propose methods to account for stratified randomization in computing standard errors, confidence intervals, and hypothesis testing.

- The linear model may include treatment by covariate interaction terms. However, when using this approach, the primary analysis can still be based on an estimate from the model of the average treatment effect (Tsiatis et al. 2008; Ye et al. 2021). As noted in the ICH E9 guidance, interaction effects may be important to assess in supportive analysis or exploratory analysis. This is because differences in treatment effects across subgroups defined by baseline covariates could be relevant to prescribers, patients, and other stakeholders and can imply that the average treatment effect gives an incomplete summary of efficacy.

C. Nonlinear Models

- Covariate adjustment with nonlinear models is often used in the analysis of clinical trial data when the primary outcome of interest is not measured on a continuous scale or is right censored (e.g., binary outcome, ordinal outcome, count outcome, or time-to-event outcome). Adjustment using nonlinear models is a potentially acceptable method for analyzing these data from a clinical trial. However, there are additional issues described below that should be considered before using nonlinear models.

- In general, treatment effects may differ across subgroups. However, with some parameters such as odds ratios, even when all subgroup treatment effects are identical, this subgroup-specific conditional treatment effect can differ from the unconditional treatment effect (i.e., the effect at the population level from moving the target population from untreated to treated) (Gail et al. 1984). This is termed non-collapsibility (Agresti 2002), which is distinct from confounding and can occur despite randomization and large sample sizes. An example of non-collapsibility of the odds ratio for a hypothetical clinical trial is illustrated in Table 1 below. The unconditional odds ratio in the hypothetical target population is 4.8, which is lower than the conditional odds ratio of 8.0 in each of the biomarker-positive and biomarker-negative subgroups. In trials with time-to-event outcomes, the hazard ratio can also be non-collapsible. Unlike the odds ratio or hazard ratio, the risk difference and relative risk are collapsible.

Table 1: Non-collapsibility of the Odds Ratio in a Hypothetical Target Population

	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Biomarker-positive	50%	80.0%	33.3%	8.0
Biomarker-negative	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

- As part of the prespecification of the estimand of interest, sponsors should specify whether the treatment effect of interest in an analysis is a conditional or unconditional treatment effect.

- Cochran-Mantel-Haenszel methods (Mantel and Haenszel 1959) are acceptable for the analysis of clinical trial data with binary endpoints if there is interest in estimating a conditional treatment effect, which is assumed to be constant across subgroups defined by a covariate taking a discrete number of levels (e.g., the value 8.0 in Table 1).

- Fitting a nonlinear regression of the outcome on treatment and baseline covariates similarly attempts to estimate a conditional treatment effect. Nonlinear models extend Cochran-Mantel-Haenszel methods by allowing adjustment for continuous covariates, such as age. In nonlinear regression models (without treatment by covariate interactions), the treatment effect is assumed to be approximately constant across subgroups defined by baseline covariates in the model and can provide more individualized information than the unconditional treatment effect if the assumption holds (and not otherwise). Nonlinear models such as logistic regression or proportional hazards regression (which can include stratification of the baseline hazard) are commonly used in many clinical settings.

- Sponsors should discuss with the relevant review divisions specific proposals in a protocol or statistical analysis plan containing nonlinear regression to estimate conditional treatment effects for the primary analysis. When estimating a conditional treatment effect through nonlinear regression, the model assumptions will generally not be exactly correct, and results can be difficult to interpret if the model is misspecified and treatment effects substantially differ across subgroups. Interpretability increases with the quality of model specification. Sponsors should discuss any planned assessments of model assumptions and implications for analyses with the relevant review division.

- Sponsors can perform covariate-adjusted estimation and inference for an unconditional treatment effect (e.g., the odds ratio of 4.8 in Table 1) in the primary analysis of data from a randomized trial. The method used should provide valid inference under approximately the same minimal statistical assumptions that would be needed for unadjusted estimation in a randomized trial. With nonlinear models using a covariate-adjusted estimator for an unconditional treatment effect, sponsors can use an appropriate bootstrap method or standard error formulas justified in the statistical literature for confidence interval construction. A

variety of statistically reliable methods have been proposed in the literature for covariate adjustment with unconditional treatment effects (Colantuoni and Rosenblum 2015).

- Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with binary outcomes (e.g., Steingrimsson et al. 2017), ordinal outcomes (e.g., Díaz et al. 2016), count outcomes (e.g., Rosenblum and van der Laan 2010), and time-to-event outcomes (e.g., Tangen and Koch 1999; Lu and Tsiatis 2008). If a novel method is proposed and statistical properties are unclear, the specific proposal should be discussed with the review division.

- As an example, the following are steps for one reliable method for covariate adjustment for unconditional treatment effects with binary outcomes that produces a resulting estimator (Steingrimsson et al. 2017; Freedman 2008) termed the "standardized," "plug-in," or "g-computation" estimator:

    (1) Fit a logistic model with maximum likelihood that regresses the outcome on treatment assignments and prespecified baseline covariates. The model should include an intercept term.

    (2) For each subject, regardless of treatment group assignment, compute the model-based prediction of the probability of response under treatment using the subject's specific baseline covariates.

    (3) Estimate the average response under treatment by averaging (across all subjects in the trial) the probabilities estimated in Step 2.

    (4) For each subject, regardless of treatment group assignment, compute the model-based prediction of the probability of response under control using the subject's specific baseline covariates.

    (5) Estimate the average response under control by averaging (across all subjects in the trial) the probabilities estimated in Step 4.

    (6) The estimates of average responses rates in the two treatment groups from Steps 3 and 5 can be used to estimate an unconditional treatment effect, such as the risk difference, relative risk, or odds ratio.

- Inverse probability of treatment weighting is another reliable method of covariate adjustment for unconditional treatment effects in randomized trials (Williamson et al. 2013).

- An analysis ignoring stratified randomization is likely to overestimate standard errors and can be unduly conservative when performing inferences for an unconditional treatment effect. The Agency recommends that the standard error computation account for stratified randomization. There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (e.g., Wang et al. 2021). The statistical properties of such methods are best understood when the number of

strata is small relative to the sample size. Sponsors can propose methods to account for stratified randomization in computing standard errors, confidence intervals, and hypothesis testing.

IV. REFERENCES

Agresti, A, 2002, Categorical Data Analysis, Second Edition, New York (NY): John Wiley & Sons, Inc.

Bugni, F, IA Canay, and AM Shaikh, 2018, Inference Under Covariate-Adaptive Randomization, Journal of the American Statistical Association, 113(524):1784-1796.

Colantuoni, E and M Rosenblum, 2015, Leveraging Prognostic Baseline Variables to Gain Precision in Randomized Trials, Statistics in Medicine, 34(18): 2602–2617..

Díaz, I, E Colantuoni, and M Rosenblum, 2016, Enhanced Precision in the Analysis of Randomized Trials with Ordinal Outcomes, Biometrics, 72(2):422-431.

Efron, B and RJ Tibshirani, 1993, An Introduction to the Bootstrap, Boca Raton (FL): Chapman & Hall.

Freedman DA, 2008, Randomization Does Not Justify Logistic Regression, Statistical Science, 23(2):237-249.

Gail, MH, S Wieand, and S Piantadosi, 1984, Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates, Biometrika, 71(3):431-444.

Lin W, 2013, Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique, Annals of Applied Statistics, 7(1):295-318.

Lu, X and AA Tsiatis, 2008, Improving the Efficiency of the Log-Rank Test Using Auxiliary Covariates, Biometrika, 95(3):679-694.

Mantel, N and W Haenszel, 1959, Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease, Journal of the National Cancer Institute, 22(4):719-748.

National Research Council, 2010, The Prevention and Treatment of Missing Data in Clinical Trials, Washington (DC): The National Academies Press.

Rosenbaum PR, 2002, Covariance Adjustment in Randomized Experiments and Observational Studies. Statistical Science, 17(3):286-327.

Rosenblum, M and MJ van der Laan, 2009, Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Hypothesis Tests Despite Incorrectly Specified Models, Biometrics, 65(3):937-945.

Rosenblum, M and MJ van der Laan, 2010, Simple, Efficient Estimators of Treatment Effects in Randomized Trials Using Generalized Linear Models to Leverage Baseline Variables, International Journal of Biostatisitcs, 6(1):13.

Steingrimsson, JA, DF Hanley, and M Rosenblum, 2017, Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions, Contemporary Clinical Trials, 54:18-24.

Tangen, CM and GG Koch, 1999, Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon Scores and Survival-Rate Estimation in a Randomized Clinical Trial, Journal of Biopharmaceutical Statistics, 9(2):307-338.

Tsiatis, AA, M Davidian, M Zhang, and X Lu, 2008, Covariate Adjustment for Two-Sample Treatment Comparisons in Randomized Trials: A Principled Yet Flexible Approach, Statistics in Medicine, 27(23):4658-4677.

Wang, B, R Susukida, R Mojtabai, M Amin-Esmaeili, and M Rosemblum, 2021. Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment, Journal of the American Statistical Association, doi: 10.1080/01621459.2021.1981338.

Williamson, EJ, A Forbes, and IR White, 2013, Variance Reduction in Randomised Trials by Inverse Probability of Treatment Weighting Using the Propensity Score, Statistics in Medicine, 33(5):721-737.

Ye, T, Y Yi, and J Shao, 2021, Inference on The Average Treatment Effect Under Minimization and Other Covariate-Adaptive Randomization Methods, Biometrika, 109(1)33-47.

Ye, T, J Shao, Y Yi, and Q Zhao, 2022, Toward better practice Of Covariate Adjustment In Analyzing Randomized Clinical Trials, Journal of the American Statistical Association, doi: 10.1080/01621459.2022.2049278.